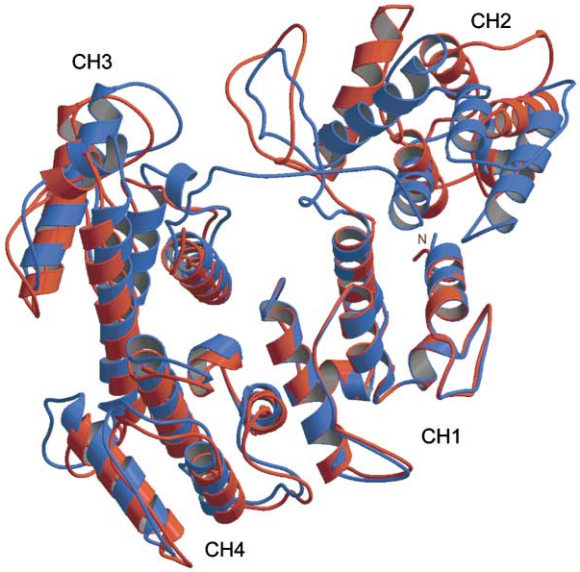

Figure 1. Structural Plasticity of the CH Domains

The structural plasticity of the CH domains is suggested in a comparison between the two molecules of the fimbrin actin binding core found in the asymmetric crystal unit (Klein et al., 2004). Each actin binding domain (ABD) contains two CH domains. The arrangement of CH3 and CH4 in ABD2 is relatively fixed between the two copies of the molecule, but the large rotation (~50°) of CH2 in this superposition shows the flexibility of the domain-domain contacts in ABD1. Image courtesy of Michael Klein.

structures of flexible tandem CH domains may not directly show the conformation of these ABDs when complexed with F-actin, but that these high-resolution structures are essential for making sense of the biochemical and genetic data about residues in both actin and actin binding proteins that are important for interactions. The key tasks now involve generating higher resolution reconstructions of F-actin decorated with these actin binding proteins, accounting for the multiple modes of binding that may be adopted, and understanding how the potential structural polymorphism is controlled within the cellular complexes.

Edward H. Egelman
Department of Biochemistry
and Molecular Genetics
University of Virginia Health Sciences Center
Jordan Hall Box 800733
Charlottesville, Virginia 22908

Selected Reading

Galkin, V.E., Orlova, A., VanLoock, M.S., and Egelman, E.H. (2003). J. Mol. Biol. *331*, 967–972.

Galkin, V.E., Orlova, A., VanLoock, M.S., Rybakova, I.N., Ervasti, J.M., and Egelman, E.H. (2002). J. Cell Biol. *157*, 243–251.

Garcia-Alvarez, B., Bobkov, A., Sonnenberg, A., and de Pereda, J.M. (2003). Structure *11*, 615–625.

Gimona, M., and Mital, R. (1998). J. Cell Sci. *111*, 1813–1821.

Goldsmith, S.C., Pokala, N., Shen, W., Fedorov, A.A., Matsudaira, P., and Almo, S.C. (1997). Nat. Struct. Biol. *4*, 708–712.

Hanein, D., Volkmann, N., Goldsmith, S., Michon, A.M., Lehman, W., Craig, R., DeRosier, D., Almo, S., and Matsudaira, P. (1998). Nat. Struct. Biol. *5*, 787–792.

Keep, N.H., Winder, S.J., Moores, C.A., Walke, S., Norwood, F.L., and Kendrick-Jones, J. (1999). Struct. Fold. Des. *7*, 1539–1546.

Klein, M.G., Shi, W., Ramagopal, U., Tseng, Y., Wirtz, D., Kovar, D.R., Staiger, C.J., and Almo, S.C. (2004). Structure *12*, this issue, 993–1013.

Liu, J., Taylor, D.W., and Taylor, K.A. (2004). J. Mol. Biol. *338*, 115–125.

Matsudaira, P. (1991). Trends Biochem. Sci. *16*, 87–92.

Moores, C.A., Keep, N.H., and Kendrick-Jones, J. (2000). J. Mol. Biol. *297*, 465–480.

Norwood, F.L., Sutherland-Smith, A.J., Keep, N.H., and Kendrick-Jones, J. (2000). Struct. Fold. Des. *8*, 481–491.

Sutherland-Smith, A.J., Moores, C.A., Norwood, F.L., Hatch, V., Craig, R., Kendrick-Jones, J., and Lehman, W. (2003). J. Mol. Biol. *329*, 15–33.

Structure, Vol. 12, June, 2004, DOI 10.1016/j.str.2004.05.003

Protein-Protein Interfaces Are Special

In this issue of *Structure*, characteristic features that distinguish protein-protein interfaces from noninterface protein surface (Halperin et al., 2004) provide insights into fundamental properties of protein recognition and contribute to improvement of docking methodologies.

Studies of protein-protein interactions are an important direction in computational structural biology. The increasing availability of crystal structures of protein complexes has allowed characterization of the interfaces between the proteins. Databases of cocrystallized protein-protein complexes are used to study the interface properties and derive relevant principles. Such currently established principles involve: (i) importance of steric complementarity, hydrophobicity, as well as electrostatic and hydrogen bonding complementarity, (ii) similarity of residue-residue and atom-atom preferences in protein-protein interfaces and protein cores, (iii) existence of "hot spots" (side chains at the interface that play a significant role in the binding), and (iv) evolutionary conservation of the interface residues.

The paper from Nussinov's group (Halperin et al., 2004), which appears in this issue, describes a systematic study of protein-protein interfaces based on a com-

prehensive database of protein complexes. For a number of years, Nussinov, Wolfson, and coworkers have been making important contributions to the development of modern approaches to the prediction of structures of protein complexes (protein docking). At the same time, they are actively involved in studies of fundamental principles of molecular recognition and their application to complex formation and protein folding. Their first nonredundant database of cocrystallized protein-protein complexes (Tsai et al., 1996) provided important information on the structure and properties of the binding sites. Their current study is based on the vastly expanded dataset of structures, which allows deep insights into the structural and physicochemical organization of protein interfaces. The results reveal correlation of hot spots and conserved residues, important variations of packing density within the interface area, and lower than expected occurrence of charge residue couples. The implications for docking methodologies involve a new outlook on the relative importance of physicochemical recognition factors and an important step in the design of docking constraints based on evolutionary considerations.

Such improved knowledge of protein recognition principles is urgently needed for further development of protein docking approaches. Adequate computational techniques to model protein interactions are important because of the growing number of known protein 3D structures, particularly in the context of structural genomics (Sali et al., 2003). The number of protein-protein interactions is significantly larger than the number of individual proteins. Protein docking techniques offer tools for 3D modeling of these interactions.

Computational structural approaches to molecular recognition were introduced in early seventies by Scheraga and coworkers (Platzer et al., 1972) for small ligand interactions with proteins. Protein-protein docking techniques were pioneered in 1978 by Wodak and Janin (Wodak and Janin, 1978) and Greer and Bush (Greer and Bush, 1978). Since then the field has grown substantially, especially starting from early nineties, through the development of powerful docking algorithms, rapid progress in computer hardware, and significant expansion of available experimental data on structures of protein-protein complexes.

In living organisms proteins recognize their partners among many other proteins and bind in a specific way in short physiological timeframes. Given the complexity of the system, the formation of a protein-protein complex is a remarkable event, based on the nature's super-efficient "energy minimization protocol" and guided by strong long- and short-range recognition factors. Modern methods of protein docking are based on our efforts to simulate and navigate the intermolecular energy landscape, and on our current understanding of the recognition factors governing complex formation.

In modeling of individual protein structures, the template-based techniques (homology and threading) have become the major driving force. Compared with ab initio protein structure predictions, the template-based approaches provide a significantly higher accuracy (Moult et al., 2003). The current situation in docking, however, is different because of two major factors. First, the structure of the complex is generally more difficult to obtain by experimental techniques (e.g., X-ray crystallography or NMR) than the structure of individual proteins. The second factor (related to the first) is that it is widely believed that the majority of functional protein-protein interactions are transient, and thus do not form complexes stable enough for crystallization. Thus, the pool of protein-protein structural templates is limited and heavily biased toward multisubunit proteins. The significance of template-based modeling of protein-protein complexes is growing, especially in such important applications as predictions of the existence of an interaction (Lu et al., 2002). For the prediction of protein-protein docking modes, however, the docking techniques are virtually exclusively ab initio ones (Janin et al., 2003). Thus, the constraints described in the Nussinov and coworkers report are of great importance.

Following recent dramatic progress in genomics, accompanied by advances in structural and computational biology, the importance of modeling protein-protein interactions has grown significantly. Accordingly, the visibility of protein-protein docking field has increased and the protein docking community has begun to organize and actively develop community-wide activities. At the First Conference on Modeling of Protein Interactions in Genomes at Charleston, South Carolina in 2001 (Vajda et al., 2002), a number of such activities were discussed and decided upon, including CAPRI and Benchmarking community-wide experiments. These activities were further developed at the Second Conference at Stony Brook, New York in 2003, CAPRI meeting at La Londe-des-Maures, France, 2002 (Janin et al., 2003), and other meetings. Comparison of the agendas of the meetings separated only by a two-year span shows dramatic progress in the studies specifically aimed at identification of the protein binding sites as a prerequisite for constrained docking or for proteins with unknown docking partners.

The future of protein docking requires new ideas and better understanding of underlying principles. The major challenge for new docking approaches is provided by the genomics era. With a rapid progress in experimental structural determination of proteins, currently about one third of individual protein structures can be modeled by relatively accurate template-based techniques (Sali et al., 2003). This percentage is expected to grow significantly in the near future. At the same time, new experimental and computational techniques yield genome-wide maps of protein-protein interactions with increasingly greater precision. The combination of these two factors paves the way for future genome-wide structural modeling of protein-protein interactions. Such modeling will reveal deep insights into fundamental principles of life at the molecular level. To become practical, it will require more accurate methods of building genome-wide maps of protein-protein interactions and the development of advanced high-throughput docking/modeling approaches.

Ilya A. Vakser
Bioinformatics Laboratory
Department of Applied Mathematics and Statistics
State University of New York at Stony Brook
Stony Brook, New York 11794

Selected Reading

Greer, J., and Bush, B.L. (1978). Proc. Natl. Acad. Sci. USA *75*, 303–307.

Halperin, I., Wolfson, H.J., and Nussinov, R. (2004). Structure *12*, this issue, 1027–1038.

Janin, J., Henrick, K., Moult, J., Ten Eyck, L., Sternberg, M.J.E., Vajda, S., Vakser, I., and Wodak, S.J. (2003). Proteins *52*, 2–9.

Lu, L., Lu, H., and Skolnick, J. (2002). Proteins *49*, 350–364.

Moult, J., Fidelis, K., Zemla, A., and Hubbard, T. (2003). Proteins *53*, 334–339.

Platzer, K.E.B., Momany, F.A., and Scheraga, H.A. (1972). Int. J. Pept. Prot. Res. *4*, 187–200.

Sali, A., Glaeser, R., Earnest, T., and Baumeister, W. (2003). Nature *422*, 216–225.

Tsai, C.-J., Lin, S.L., Wolfson, H.J., and Nussinov, R. (1996). J. Mol. Biol. *260*, 604–620.

Vajda, S., Vakser, I.A., Sternberg, M.J.E., and Janin, J. (2002). Proteins *47*, 444–446.

Wodak, S.J., and Janin, J. (1978). J. Mol. Biol. *124*, 323–342.

Structure, Vol. 12, June, 2004, DOI 10.1016/j.str.2004.05.002

The Protein Surface Is a Moving Target

A clear wake-up call that we need to improve the crystallographic treatment of partial disorder is provided by DePristo, de Bakker, and Blundell (2004) in the May issue of *Structure*.

This paper by DePristo et al. (2004) is a watershed treatment of a problem that has been recognized in some earlier work (Ohlendorf, 1994; Mowbray et al., 1999; Kleywegt, 1999; Vitkup et al., 2002), but never before analyzed in a sufficiently simple, general, and unbiased form to demand the attention both of practicing crystallographers and of the general community using such structures. The dynamic nature of protein structures, especially pronounced at the molecular surface, has long been recognized from theoretical calculations, from NMR and other spectroscopic data, and even from crystallography itself, as epitomized by the recent 0.66 Å resolution structure of aldose reductase (Howard et al., 2004), where 99 of the 314 side chains and substantial regions of backbone show alternate conformations.

For protein structures at more typical resolution, however, this recognition produces a difficult dilemma: the dynamic side chains simply disappear or become uninterpretable, and the data are definitely not sufficient to justify fitting multiple conformations. Sometimes such atoms are omitted, which is a reasonable decision but causes difficulties for users. Most often the crystallographer (or, in some cases now, the automated routine) chooses a single conformation that intersects at least some small peak of density and refines from there. With luck, the result may sometimes match one of the real conformations, but it never represents the real ensemble.

To deal with this problem, DePristo et al. use their RAPPER program (DePristo et al., 2003) to generate an ensemble of starting models close to the PDB structure, within positive electron density, and obeying bond length, bond angle, Ramachandran, and rotamer criteria. Those models are then refined by quite standard methods, and a set of five are chosen in each case that show global quality criteria (R, R_{free}) and global averages of local criteria (geometry, all-atom clashes, and percentage of Ramachandran or rotamer outliers) at least as good as for the original PDB coordinates. Usually the ensemble R factors were found to be lower than those for either the individual or the original models. Thus, each of these alternative models is an equally good representation of the structure. Yet they differ quite substantially from one another, especially on the surface and at lower resolutions. It is difficult to avoid the conclusion that accuracy is lower and heterogeneity greater than is usually estimated.

Several examples are shown in Figure 1. (A) shows Asp 1 in 1AAC at 1.3 Å resolution, where the electron density ends in a quite rounded blob but only a single conformation was originally fit. (B) shows Arg 41 in 1G35 at 1.8 Å, where the electron density almost disappears beyond $C\gamma$ and there surely must be more than two different conformations. I would argue that those multiple conformations must all be rotameric in such cases, where there are no interactions that could hold the side chain out of its local energy minima. (C), from 9ILB at 2.3 Å resolution, is a reminder that large sections of these structures are extremely well-ordered and accurately determined.

The take-home lessons from this work are that there is a greater degree of heterogeneity and inaccuracy in protein crystal structures than usually acknowledged, and that the crystallographic community would benefit from adopting methods better able to recognize and represent that heterogeneity even at moderate resolutions. The present version of RAPPER is well equipped to demonstrate and even quantify this problem, but is not yet ready for routine use in refinement because it produces occasional residues with unreasonable backbone or sidechain conformations. We look forward, however, to either an updated RAPPER or some similar routine that can support a more explicit treatment of protein conformational heterogeneity.

Jane S. Richardson
Department of Biochemistry
Duke University
Durham, North Carolina 27710